

The Prediction of Customer Churn

October 2022

Florecita Prameswari



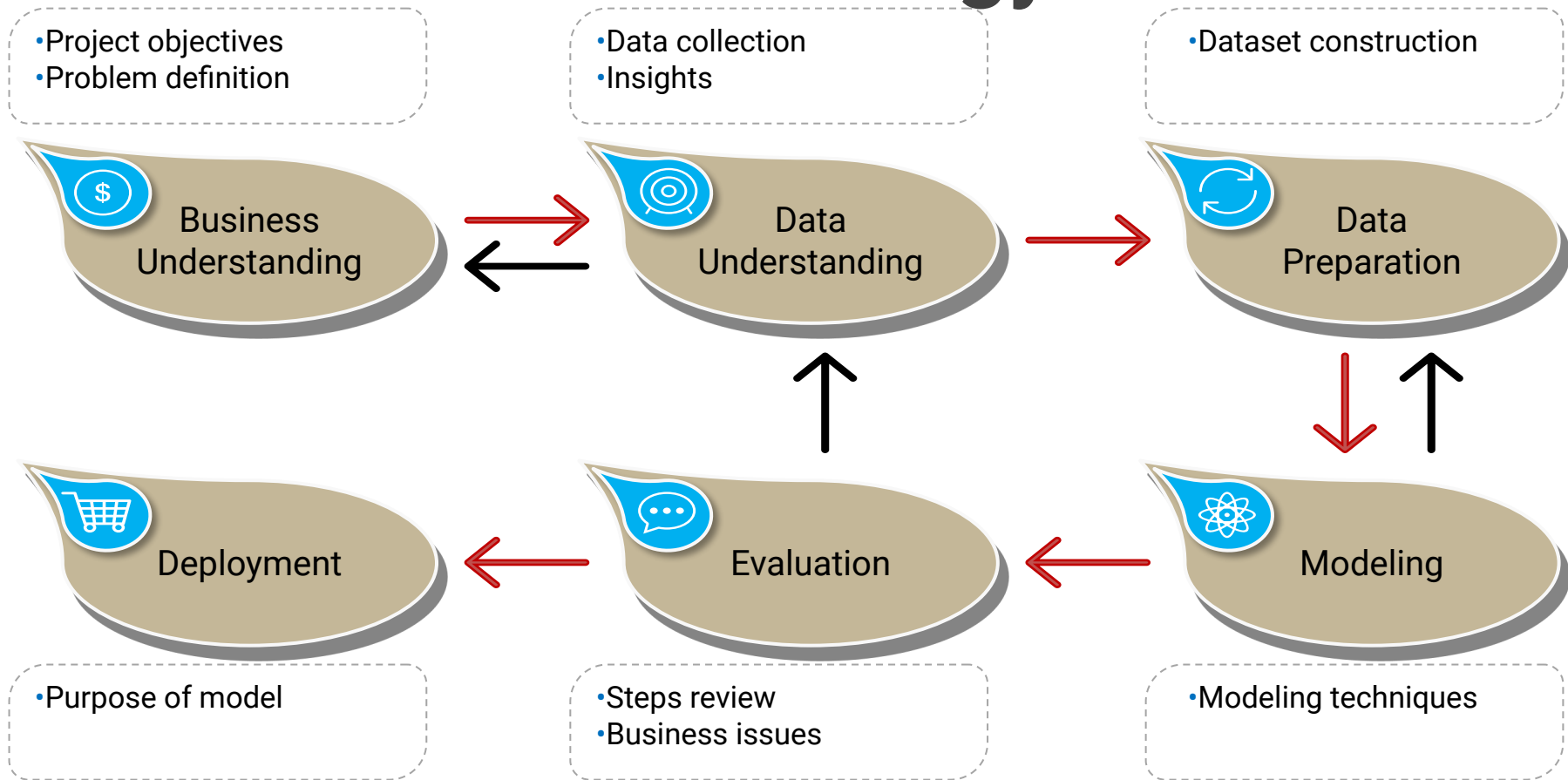
Problem definition

Customer churn is important because every time a client **leaves**, it represents a significant investment **lost**. Both time and effort need to be channelled into replacing them.

Goals

To take effective action to retain the customer before it is too late by being able to predict when a client is likely to **leave**

Methodology



Exploratory Data Analysis

- Data Quality Checking
- Descriptive Statistics
- Target Variable Distribution (Train and Test)
- Numerical Feature Distribution (Train and Test)
- Categorical Feature Distribution (Train and Test)

Exploratory Data Analysis

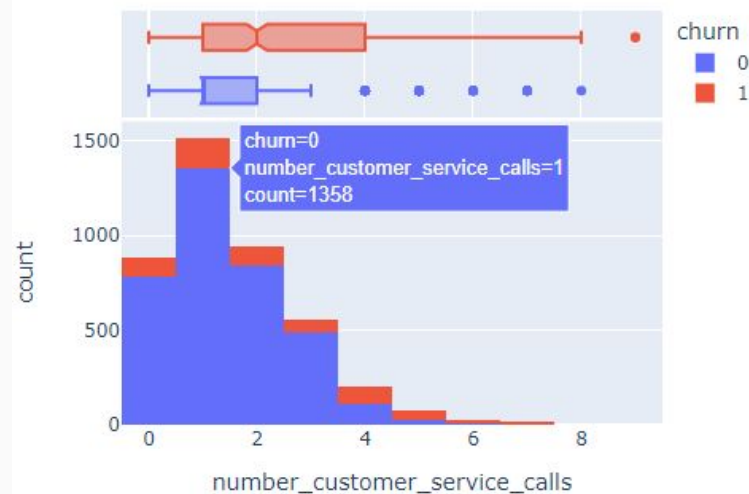
Based on analysis data, there are so many customer who still subscribed or will likely no to churn used the customer services

It means we are already in a good track by providing customer call services to our customer



Churn (0) : Still Subscribed

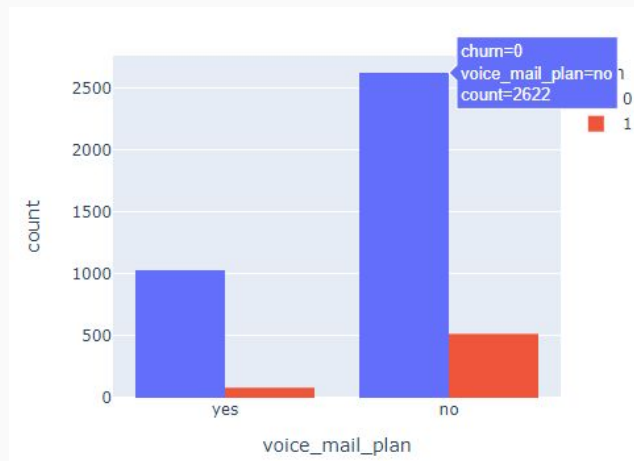
Churn (1) : Not Subscribe



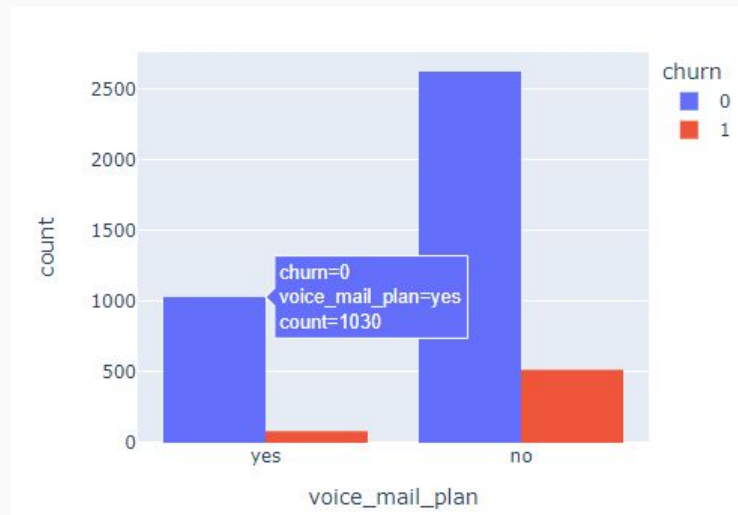
Exploratory Data Analysis

Based on analysis data, there are so many subscribed customer who did not have voicemail plan yet. Understanding the voicemail plan with a effective and cost-efficient plan can help forecast subscribed customer needs

**Churn (1) : Not
Subscribe**

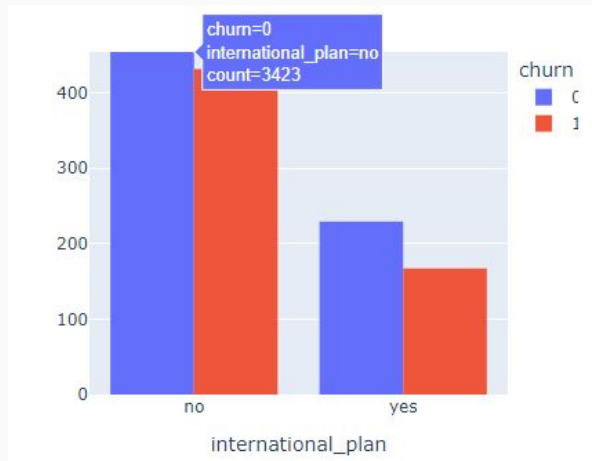


**Churn (0) : Still
Subscribed**



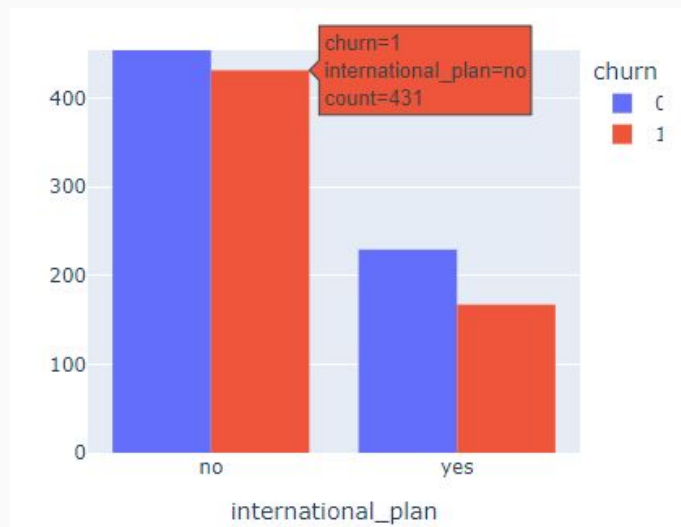
Exploratory Data Analysis

Based on analysis data, there are many customer who churn had a international plan and there are many customer who is not churn also did not have the international plan.



Churn (0) : Still Subscribed

Churn (1) : Not Subscribe



Data Preprocessing

Data Cleaning

Fill in missing values

Delete duplicate data

Feature Encoding

One-hot Encoding to convert numerical categorical variables into binary vectors

Normalization

Use Min-Max Scaler to transform numerical data value by scaling into range between 0 - 1

Result of Support Vector Machine Model

Prediction : is there any chance that the customer will likely to churn?

True Positive

Interpretation: You predicted customer will churn and it's true.

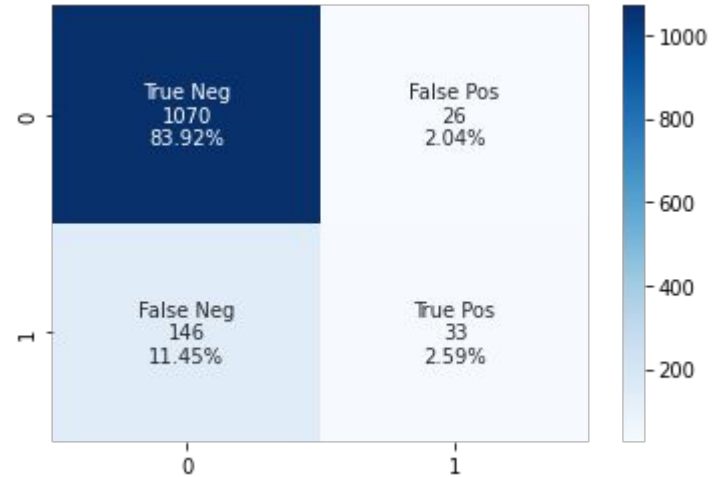
Based on data there is only 2.59% chance that the customer will likely to churn.

True Negative

Interpretation: You predicted customer will not churn and it's true.

Based on data there is 83.92% chance that the customer will likely not to churn.

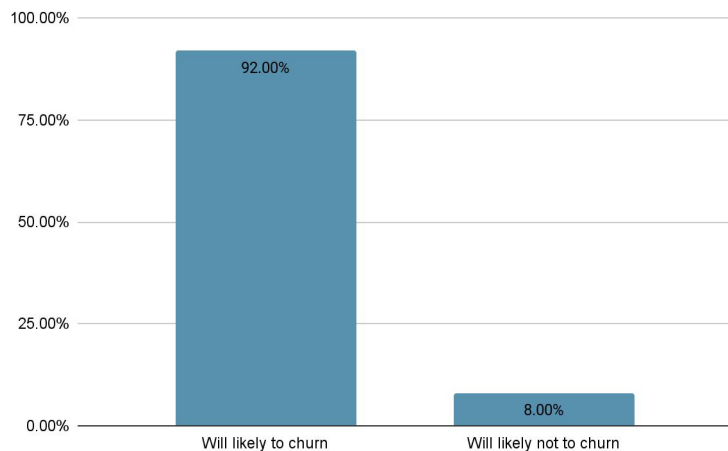
Prediction score on data testing 92%
Prediction score after undersampling 58%



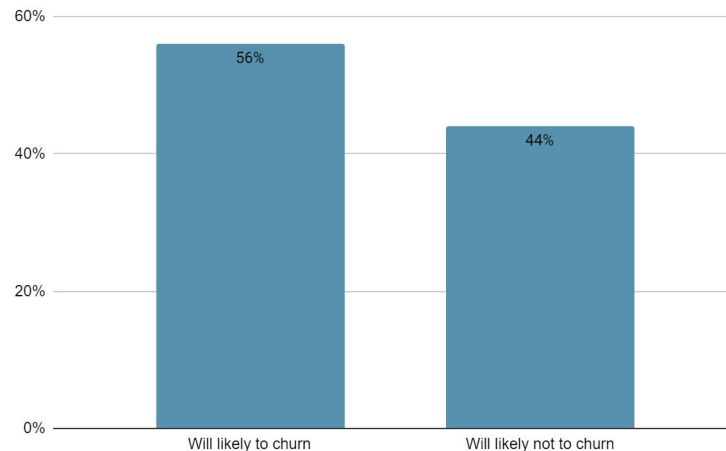
Confusion matrix SVM model with
86.5% accuracy

Case Interpretation

Support Vector Machine



Support Vector Machine Undersampling



Result of K-Nearest Neighbors Model

Prediction : is there any chance that the customer will likely to churn?

True Positive

Interpretation: You predicted customer will churn and it's true.

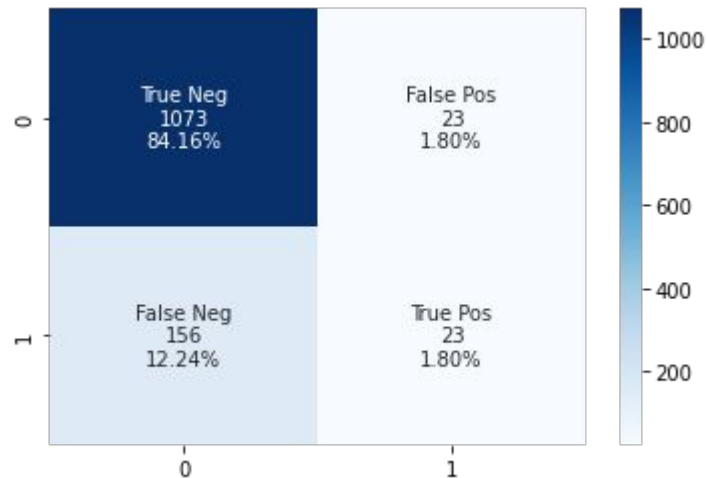
Based on data there is only 1.80% chance that the customer will likely to churn.

True Negative

Interpretation: You predicted customer will not churn and it's true.

Based on data there is 84.16% chance that the customer will likely not to churn.

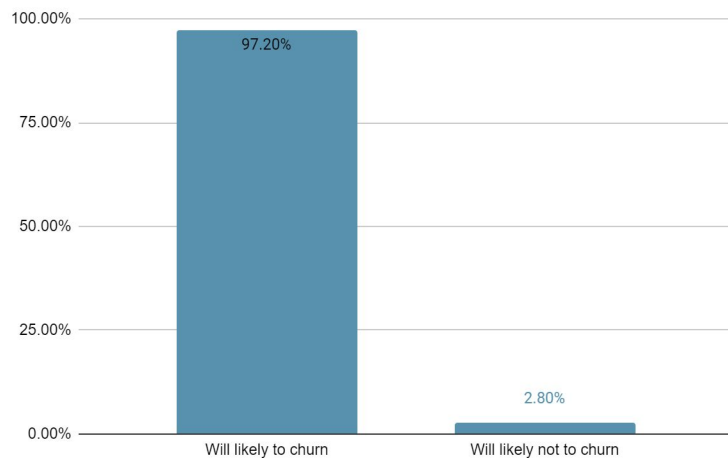
Prediction score on data testing 92%
Prediction score after undersampling 63%



Confusion matrix KNN model with
86.5% accuracy

Case Interpretation

K-Nearest Neighbors



K-Nearest Neighbors Undersampling

